

# Multiscale Sequential Convolutional Neural Networks for Simultaneous Detection of Fovea and Optic Disc

Baidaa Al-Bander<sup>1,3</sup>, Waleed Al-Nuaimy<sup>1</sup>, Bryan M. Williams<sup>2</sup>, Yalin Zheng<sup>2</sup>

<sup>1</sup>Department of Electrical Eng. and Electronics, University of Liverpool, Liverpool L69 3GJ, UK

<sup>2</sup>Department of Eye and Vision Science, University of Liverpool, Liverpool L7 8TX, UK

<sup>3</sup>Department of Computer and Software Engineering, University of Diyala, Iraq

Email: {hsbalban, wax, bryan, yzheng} @liverpool.ac.uk

**Abstract-** Detecting the locations of the optic disc and fovea is a crucial task towards developing automatic diagnosis and screening tools for retinal disease. We propose to address this challenging problem by investigating the potential of applying deep learning techniques to this field. In the proposed method, simultaneous detection of the centers of the fovea and the optic disc (OD) from color fundus images is considered as a regression problem. A deep multiscale sequential convolutional neural network (CNN) is designed and trained. The publically available MESSIDOR and Kaggle datasets are used to train the network and evaluate its performance. The centers of the fovea and the OD in each image were marked by expert graders as the ground truth. The proposed method achieves an accuracy of 97%, 96.7% for the detection of the OD center and 96.6%, 95.6% for the detection of the foveal center of the MESSIDOR and Kaggle test sets respectively. Our promising results demonstrate the excellent performance of the proposed CNNs in simultaneously detecting the centers of both the fovea and OD without human intervention or handcrafted features. Moreover, we can localize the landmarks of an image in 0.007 seconds. This approach could be used as a crucial part of automated diagnosis systems for better management of eye disease.

**Keywords—** Diabetes; Fovea Detection; Optic Disc Detection; Convolutional Neural Networks.

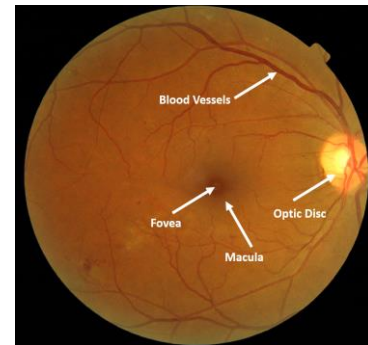
## 1. INTRODUCTION

The knowledge of the optic disc (OD) and fovea (macula center) locations in the retina is considered essential for the diagnosis and screening of many retinal diseases, such as glaucoma, diabetic maculopathy (DM) and age-related macular degeneration (AMD). The significance of detecting the fovea is that the closer a lesion is to it, the more likely the lesion is to cause visual impairment or blindness. On the other hand, the OD center is often regarded as a reference point for locating other retinal structures. For example, it can be used as the starting point for tracking retinal vessels in blood vessel tracking algorithms [1]. In addition, the OD diameter ( $\rho$ ) is usually used as the reference to measure the size and location of other anatomical and pathological structures in the retina. On average the vertical OD diameter is about 1800 $\mu$ m.

The OD appears as a bright yellowish oval region within color fundus images through which the blood vessels enter the eye. The macula is the center of the retina which is responsible for our central vision. The fovea is a small depression in the center of the macula. It has a darker appearance compared to the surrounding retinal tissue due to the high concentration of macular pigment. **Fig. 1** shows a color retinal fundus image with the key anatomical structures denoted. The location of the

fovea center is about  $2.5\rho$  from the optic disc center. The foveal radius is between  $1/3$  and  $1/4$  of the macula radius which is roughly equal to one optic disc diameter ( $\rho$ ) [2] [3].

Recently, the automatic localization and detection of retinal anatomical structures from digital fundus images has received increasing attention in the medical image processing community [19]- [22]. This may support the development of computer aided diagnosis (CAD) tools for the better management of eye disease. Despite considerable effort in this field, the problem of localizing the centers of the OD and the fovea remains unsolved in retinal fundus image analysis.



**Fig. 1** An example fundus image illustrating the key retinal anatomical structures. Note the darker appearance at the fovea and blood vessels originating at the optic disc.

In this paper, a multiscale sequential deep learning technique is proposed which is aimed at detecting the centers of the OD and the fovea. The main contributions and advantages of this work are summarized as follows:

In this paper, a deep learning technique is proposed which is aimed at detecting the centers of the OD and the fovea. The main contributions and advantages of this work are summarized as follows:

1. The application of deep convolutional neural networks to the detection of retinal landmarks is novel and promising. We develop a suitable convolutional neural network to detect specifically the optic disc and fovea centres.
  - a. **Speed and automation:** This results in a fast method requiring no user input.
  - b. **Independence:** The method is not dependent on other techniques succeeding such as segmentation or detecting other landmarks.
  - c. **No handcrafted features:** Since features do not need to be manually defined, we avoid the difficulty

encountered by conventional machine learning algorithms in identifying the best feature set that represents the data. This also removes the requirement of a skilled technician to identify such features manually which takes a considerable amount of time and can produce subjective results, particularly with a large dataset.

- d. **Accurate simultaneous detection:** We detect more than one position simultaneously, retaining high accuracy for each.
  - e. **Robustness:** The method is robust in the sense that it continues to work well even on poor quality images.
2. We develop a multiscale approach to convolutional neural networks to focus on the region of interest.
    - a. **Improved Accuracy:** This approach allows the method to focus on the region of interest, removing redundant background data from consideration and facilitating refinement of the localisation. This results in significantly increased accuracy in the cases of the fovea and the optic disc.
  3. Inter-dataset training and evaluation using multiple datasets.
    - a. **Generalisation:** This demonstrates generalisation of the method to new data, from separate datasets and graders, and captured from different devices.
  4. We incorporate variable optic disc radius (R) into evaluation criteria.
    - a. **Evaluation accuracy:** Incorporating this variable measure into our testing allows more accurate evaluation while others' use fixed R value for evaluation.

The remainder of this paper is organized as follows. Section 2 provides a brief review of the previous work related to the detection of the OD and the fovea. Section 3 describes the proposed methodology for detecting the OD and fovea locations. The experiments and results are described in Section 4. This work is discussed in Section 5 and the paper is concluded in Section 6.

## 2. RELATED WORK

In the literature, there has been a number of studies conducted to determine the locations of the fovea and OD. Many of these studies only locate either the OD or fovea and not both. Below is a brief review of the major algorithms published in the literature for detecting the OD, followed by fovea detection methods.

Many of the reported methods use geometric information of the vascular tree to detect the OD [4]-[8]. Hoover and Goldbaum [4] exploited the spatial relationship between the OD and retinal blood vessels and proposed a fuzzy convergence algorithm to locate the origination point of the blood vessel network. This origination point was considered as the OD center in the retinal fundus image. Foracchia *et al.* [5] proposed a geometrical model to calculate the general direction of retinal blood vessels at any given location in an image using the coordinates of the OD center as the two model parameters. The simulated annealing optimization technique was used to

identify these two parameters. Furthermore, Fleming *et al.* [6] presented a method based on the elliptical form of retinal blood vessels to obtain the approximate locations of the OD and fovea. The circular edge of the OD and the darker appearance of the fovea were exploited to refine these approximated locations. In addition, Tobin *et al.* [7] used accurate vasculature segmentation results for optic disc detection by determining density, average thickness, and average orientation of the blood vessels in relation to the position of the OD. Youssif *et al.* [8] described a method that can detect the optimal OD center point by measuring the difference between the matched filter output and the vessels' directions.

Niemeijer *et al.* [9] formulated the problem of detecting the OD and foveal centers as a regression problem. They utilized a kNN regressor to measure the distance in an image to the object of interest at any given location using a set of features extracted at that location. Furthermore, a method based on Sobel operators and the Hough transform for the detection of the OD in retinal fundus images was formulated by Zhu *et al.* [10]. They determined the center and radius of the OD by approximating the margin of the optic nerve head into a circle using the Hough transform. Moreover, Lu *et al.* [11] designed a technique based on the circular transformation to locate the circular shape of the optic disc and color variation across the OD boundary. The center and the boundary of the optic disc were located by exploiting the pixels with the maximum variation along radial line segments.

Yu *et al.* [12] presented a method for detecting the optic disc location using template matching techniques. The OD location was determined using the characteristics of the vessels on the OD. In [13], Dehghani *et al.* proposed a histogram based method which uses four images from the DRIVE dataset as a template to locate the center of the OD where each histogram represents one color from the RGB color image components (red, blue, and green). The template was constructed by calculating the average of these histograms. Harangi *et al.* [14] adapted the most recent OD detectors and organized them into an ensemble and complex framework in order to merge their strengths and maximize the accuracy of OD detection. To determine the final OD position, a maximum-weighted clique was founded. Recently, Calimeri *et al.* [15] have presented a method based on fine-tuned convolutional neural network to localize the OD location.

Many of the fovea localization approaches presented in the literature have exploited the vasculature and other contextual information. Li and Chutatape [16] presented a model-based approach by combining the information provided by the main vessel arcades and the low intensity pixels in the fovea region. A parabola fitting method was used to detect the fovea and the fovea center was identified using a thresholding scheme in the region of interest.

Niemeijer *et al.* [17] formulated a method based on a cost function that is based on both global and local cues to find the fovea. In addition, mathematical morphology and anatomical knowledge based methods were used to estimate the location of the fovea by Welfer *et al.* [18]. In their proposed system, extracting the region of interest containing the fovea was

achieved initially by calculating the center and diameter of the OD. After that, a set of fovea candidates was obtained using a morphological operation. To detect the center of the fovea, it was selected as the centroid of the darkest candidate.

Qureshi *et al.* [19] proposed a method based on a combination of several algorithms for detecting the fovea and OD. They proved that ensemble algorithms can achieve better performance than a single algorithm for detecting these centers. Moreover, a fast radial symmetry transform was used by Giachetti *et al.* [20] for the detection of the fovea and OD centers. The centers of symmetry of dark and bright regions were detected by applying the transform on coarsened and vessel-inpainted images and the results were combined with a vascular density estimator.

Gegundez-Arias *et al.* [21] detected the location of the fovea center by means of prior known anatomical features. These features were used to localize a ROI fovea-containing sub-image. A multi-thresholding scheme using gray-level value criteria was applied and a contour map was created to calculate the fovea center. In [22], Aquino *et al.* formulated a method based on combining the visual and anatomical features of the macula and the OD for detecting the fovea center.

From the above review, it can be noticed that most of the previous studies have exploited the visual appearance or anatomical features for the detection of the OD and fovea in order to identify their positions [4], [6], [11], [7], [18], and [20-22]. These methods will suffer when these features are very weak or invisible due to pathologies. Some other methods rely on machine learning algorithms and feature extraction to localize and detect anatomical structures [9], [14], and [19], but the accuracy of these methods largely depends on the type and quality of the feature sets which are hand-crafted. Inspired by our observations, we propose to introduce new deep learning techniques to address this.

The main aim of the proposed method is to develop a deep learning based approach to simultaneously detect both the OD and the fovea locations. Based on deep convolutional neural networks (CNNs), our new approach is expected to be independent of the manual detection of anatomical features of retinal landmarks. Moreover, in contrast to more traditional machine learning and feature extraction algorithms, the hierarchically extracted features are automatically learned from data and not designed manually. In addition, the proposed approach has yielded promising results and outperforms conventional neural networks, which demonstrates that deep learning techniques will be able to support robust and accurate detection of the OD and foveal centers.

### 3. MATERIALS AND METHODS

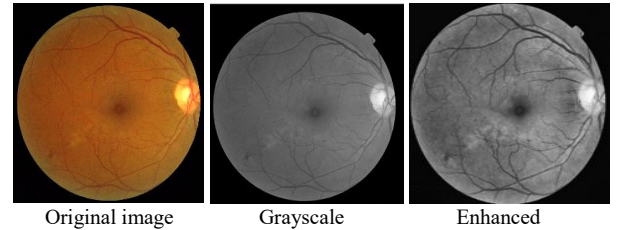
#### A. Materials

The MESSIDOR [23] and Kaggle [24] datasets have been used in this work. The MESSIDOR database comprises 1200 images captured using a color 3CCD camera on a Topcon TRC NW6 with 45 degree field of view. The MESSIDOR images were captured using 8 bits per color plane at a size of 1440×960, 2240×1488 or 2304×1536 pixels. Moreover, 10,000 images of the Kaggle dataset provided by Kaggle in their diabetic

retinopathy detection competition are used for training and testing. The optic disc and foveal center point coordinates were not provided in the original dataset for both datasets, for this work they were obtained from annotations from a combination of two expert graders from the Liverpool Reading Centre. An in-house program developed in Matlab (version 2016a, Mathworks Inc, Natick, MA) was used by the grader. This software program was developed to support annotations of anatomical or pathological features required by clinical trials, and allows the grader to visualize the image, selecting the location by mouse click and make correction on the selection. These annotated locations together with the images were used to train and evaluate the performance of the implemented networks.

#### B. Pre-Processing

It is worth noting that detecting the centers of the fovea and OD is a regression task. It seems unnecessary to use color information because the colors may just add extra complexity. For this reason, all of the images were converted to grey scale for use. For the purpose of this study, the images were resized to 256×256 pixels and the annotated center point coordinates of both the OD and fovea were scaled accordingly. The contrast of the resized images was enhanced by applying the contrast-limited adaptive histogram equalization technique [25] so as to reduce uneven illumination in the images as shown in **Fig. 2**. The pixel values of the enhanced images were scaled between [0, 1] and the coordinates of the center points were scaled between [-1, 1].



**Fig. 2** Image Pre-Processing stages.

#### C. Deep Convolutional Neural Network Architecture

In contrast to conventional shallow classifiers, such as neural networks and support vector machines, for which a feature extraction step is essential, hierarchies of significant features are learnt by deep learning algorithms directly from the raw input data. Recently, deep convolutional neural nets (convnets) have succeed in improving many computer vision applications such as image classification [26], object recognition [27], and keypoint localization [28]. In addition, some interesting results have been seen in biomedical applications such as neuronal membrane segmentation [29], [30] and other applications [31], [32].

A typical CNN comprises one or more convolutional layers alternated with pooling layers (subsampling layers) and then followed by fully connected layers (FC) and finally a classification/regression layer. CNNs can be considered as a special form of feedforward multilayer perceptron neural networks (MLPs). However, the number of parameters that need to be tuned is reduced to a level that becomes tractable for

the current computing power. For example, in convolutional layers, a limited number of convolutional kernels is needed.

1. *Convolutional Layer*: The convolutional layer [33] represents the core building block of a deep CNN. The neurons in the convolutional layer connect to local regions of the input and compute their outputs based only on these local regions. This layer is parameterized by a set of learnable filters (kernels) convolved over the width and height of the input image and the result of each filter is called a feature map. Given an input volume size  $N_i \times N_i \times D_i$ , the filter or receptive field size  $F$ , the depth of the convolutional layer  $K$ , the stride parameter  $S$ , and the amount of zero padding  $P_i$ , the number of neurons in the output volumes  $N_o \times N_o \times D_o$  can be calculated by the formula

$$N_o = \frac{N_i - F + 2P}{S} + 1; \quad D_o = K, \quad (1)$$

where the value of the stride parameter  $S$  should be chosen such that  $N_o$  is an integer.

2. *Max-pooling Layer*: The feature map resulting from a convolution layer is usually subsampled with  $R \times R$  non-overlapped regions (windows), where  $R$  is a hyper-parameter that can be empirically defined by the user. This window is shifted over the feature map: each time the value within this window which is most responsive (highest activation value) is selected while other values are neglected. The purpose of this layer is to speed up convergence by reducing the number of parameters and amount of computation in the deep neural network, and to provide translation invariance [34].

Given an input volume of size  $N_i \times N_i \times D_i$ , max-pooling window size  $R \times R$ , and the stride parameter  $S$ , the number of neurons in the output volumes  $N_o \times N_o \times D_o$  is calculated by the formula

$$N_o = \frac{N_i - R}{S} + 1; \quad D_o = D_i \quad (2)$$

3. *Dropout Layer*: A dropout layer [35] is an effective regularization strategy that stochastically adds noise to the hidden layers of deep neural networks. More specifically, the overfitting problem can be alleviated by randomly dropping out the output of each hidden unit with a certain probability at each training step (i.e. multiplying hidden activations by Bernoulli distributed random variables that take the value  $1/p$  with probability  $p$  and 0 otherwise;  $p = 1$  means no drop out and low values of  $p$  imply more dropout). A deactivated unit will not take part in forward propagation or backpropagation in the training stage that is achieved using the stochastic gradient descent (SGD) algorithm. At the testing stage, all of the units are re-enabled by multiplying them with one minus the probability  $p$  of masking.

4. *Fully Connected Layer*: This usually represents the final layers of a deep neural network architecture. Each node in the fully connected layer is completely connected to all of the nodes in the previous layer and the weights of these links are specific to each node. The number of neurons in the fully connected layers is considered as a hyper-parameter to be empirically chosen.

5. *Activation Functions*: a rectified function is used as an activation function for all of the layers (except for the final layer) in our implemented network. A unit employing the rectifier is called a Rectified Linear Unit (RELU). This is the

most common activation function used in deep neural networks because it is less susceptible to vanishing gradient problems [36]. The rectified function is defined by the formula:

$$\varphi: x \mapsto \max(0, x). \quad (3)$$

Since in this work a regression problem is being dealt with, a linear function as a linear combination of the activations in the fully connected layer is used in the top layer (output layer) of our network architecture.

The block diagram of the proposed deep multiscale sequential convolutional neural network is presented in Fig. 3.

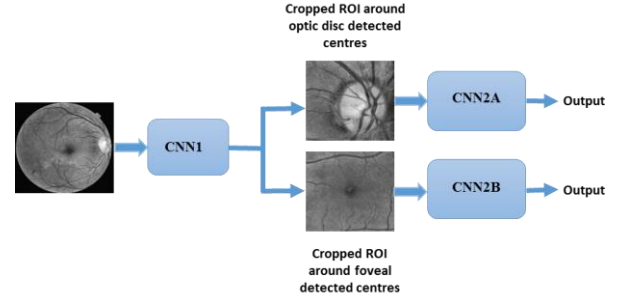


Fig. 3 Block diagram of the proposed system.

The proposed system consists of two stages, in the first stage the whole resized images along with the scaled centers are fed to the implemented CNN. The output of the first stage is the centers of both the OD and F. In the second stage, the detected centers from the first CNN are used to obtain the refined regions of interest of both OD and F by cropping the region around these centers by  $2R$  radius value ( $R$  represents OD radius). These resized ROI for both the F and OD along with the scaled ground truth centers are used to train the CNNs in the second stage. Therefore, the first stage is used to obtain the ROIs for both F and OD while the second stage is aimed to detect the centers by classifying the features extracted automatically by the convolutional filters. As we go deeper through the convolutional neural network, the convolutional layers are able to describe more and more complex features.

#### D. Performance Evaluation

In the literature, the 1R criterion (where  $R$  refers to the OD radius) is the most common criterion used to evaluate the performance of retinal landmark detection methodologies. The distance between the ground truth and the obtained location of the structure of interest (i.e. the OD or foveal center for this application) is compared with the  $R$  value in each image to determine the validity of the location determined by the automated detection methods.

In this work, both the optic disc and foveal center positions were known from expert annotations. Moreover, the location of the fovea center is about  $2.5 \rho$  from the OD center. The optic disc diameter ( $\rho$ ) and consequently the OD radius ( $R$ ) can be calculated for each image  $i$  using Equation (4).

$$\rho_i = \frac{\sqrt{(X_{ODr(i)} - X_{Fr(i)})^2 + (Y_{ODr(i)} - Y_{Fr(i)})^2}}{2.5} \quad (4)$$

Then,  $R_i = 0.5 \rho_i$  where  $X_{ODr}, Y_{ODr}$  and  $X_{Fr}, Y_{Fr}$  are the horizontal and vertical coordinates of the OD and fovea centers



respectively marked by expert graders.

#### 4. EXPERIMENTS AND RESULTS

In this work, different network architectures and data augmentation strategies were evaluated in comparison to conventional neural networks. All of the experiments were conducted on an HP Z440 running Linux Mint with 16GB RAM, an Intel Xeon E5 3.50GHz processor and NVIDIA GTX TITAN X 12GB GPU card with 3072 CUDA parallel-processing cores. The Lasagne [37] Python deep learning is used to implement and train our convolution neural networks. Built on the top of Theano [38], Lasagne has efficient implementations of each of the CNN layers, a diversity of activation functions, many optimization methods, and transparently supports training networks on GPUs.

To train the networks by updating the weights, SGD with a momentum optimization algorithm having an adaptive learning rate (start=0.03, stop=0.0001) and adaptive momentum parameter (start=0.9, stop=0.999) is used. The weights of the kernels for the implemented convolutional layers are initialized from a uniform distribution within chosen intervals. These intervals are configured by Lasagne depending on the weight initialization technique proposed in [39]. Furthermore, the objective function to be minimized is mean squared error (MSE) since we are dealing with a regression problem:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (5)$$

In order to reduce the overfitting effect, the size of the training data is increased artificially by applying data augmentation. More specifically, the training data is augmented by flipping images left to right while the annotated OD and foveal centers were flipped accordingly. As a result of this, the size of the training data has been doubled.

The deep network was trained with 1000 epochs. An early stop strategy is used so the training will stop when there is no improvement in learning or performance on the validation set starts to worsen. The early stop value was set to 100 epochs where the learning stops after 100 epochs and the best weighting values are retained if the validation error stops improving early. The architecture of CNN with the best performance is described and shown in **Table 1** and **Fig. 4**.

In **Table 1** the last column shows the size of the filters, the window size used for max-pooling, and the probability of dropping a node (Bernoulli (p)) in each layer. No zero padding and a stride of 1 pixel were used for each convolutional layer while non-overlapped pooling (stride= pool size) was used in each max-pooling layer.

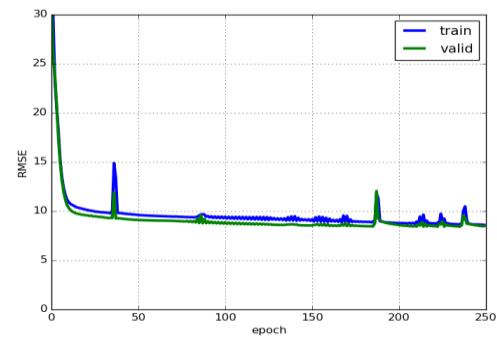
**Table. 1** Architecture of deep neural network with the best detection performance.

Name	Size	No. of outputs	No. of filters	Size of filter, max pooling, probability
input	1×256×256	65536	-	-
conv1	8×254×254	516128	8	filter size=(3,3)
conv2	8×252×252	508032	8	filter size=(3,3)
conv3	8×250×250	500000	8	filter size=(3,3)
conv4	8×248×248	492032	8	filter size=(3,3)

dropout1	-	-	-	dropout1_p=0.1
conv5	16×246×246	968256	16	filter size=(3,3)
conv6	16×244×244	952576	16	filter size=(3,3)
conv7	16×242×242	937024	16	filter size=(3,3)
pool1	16×121×121	234256	-	maxpool size=(2,2)
dropout2	-	-	-	dropout2_p=0.3
conv8	32×120×120	460800	32	filter size=(2,2)
conv9	32×119×119	453152	32	filter size=(2,2)
conv10	32×118×118	445568	32	filter size=(2,2)
pool2	32×59×59	111392	-	maxpool size=(2,2)
FC	350	350	-	-
dropout3	-	-	-	dropout3_p=0.5
FC	350	350	-	-
output	4	4	-	-

For the sake of comparison, a conventional neural network with three layers (input, hidden, output) is implemented to evaluate the effect of adding more layers in deep learning. This network is trained with 250 epochs and 200 neurons are used in the hidden layer. The size of the input layer is equal to the size of the input image and the size of the output layer is four neurons (x and y coordinates of the OD and fovea centers respectively).

Learning performance of the implemented networks is monitored during the training stage by plotting the learning curves for both training and validation sets by determining the root mean squared error (RMSE). **Fig. 5** and **Fig. 6** show the difference in terms of performance between the conventional neural network (NN) model and the deep model during the training stage. Clearly, it can be observed that the deep neural network has improved performance with much lower error than the conventional neural network model.



**Fig. 5** Performance of the conventional neural network during training. It shows that simple model suffers from an underfitting problem where the complexity of the network isn't sufficient to capture the import features of the landmarks.

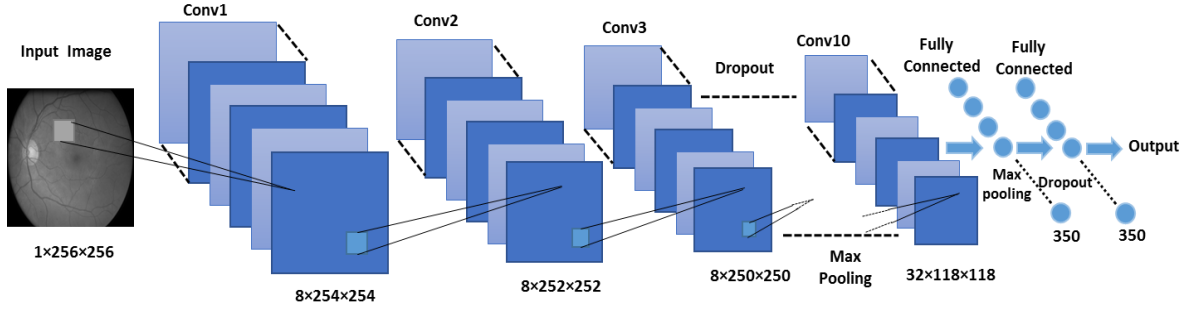


Fig. 4 Block diagram of convolutional neural network.

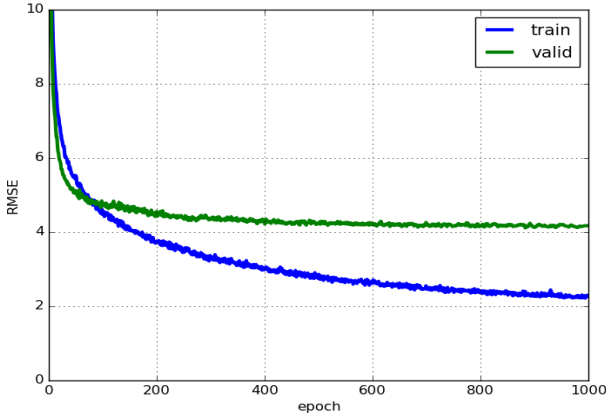


Fig. 6 Performance of the deep neural network during training. It shows that the RMSE for both training and validation data is lower than the conventional neural network with slightly overfitting and thus better landmark detection performance.



Fig. 7 Example shows 1R, 0.5R, and 0.25R of OD.

For the purpose of performance analysis of the proposed system for detecting the OD and fovea, the detection accuracy was computed as the ratio between the number of testing images with detected centers satisfying the 1R, 0.5R and 0.25R conditions (Fig. 7 explains these criteria) and the total number of testing images. In addition to the accuracy measure, the mean error (also called normalized localization error) and standard deviation are also calculated. The normalized localization error is calculated by dividing the Euclidean distance between the actual and computed OD (or foveal) centers with the  $\rho$  in each testing image. The detection performance of the neural network and deep neural network is shown in Table 2. The effect of image enhancement is also reported in Table 2 for information. In this table, the MESSIDOR dataset has been randomly divided into 70% for training and validation and the remaining 30% for testing.

Table. 2 Performance of different networks (STD: standard deviation): The networks are trained and tested on Messidor dataset (1R criterion).

Model Name	Optic Disc			Fovea		
	Acc.	Mean	STD	Acc.	Mean	STD
Simple Model (NN)	59.5	0.5683	0.5681	86.2	0.2757	0.2176
Deep model without enhancement	96.0	0.1692	0.2533	96.0	0.1320	0.1330
Deep Model	96.89	0.1596	0.2374	97.78	0.1325	0.1265

In Table 3 and Table 4, the proposed system was evaluated using the MESSIDOR and Kaggle datasets, where 7000 Kaggle images were used for training and validation (20% of training data is used as validation data) and the remaining 3000 Kaggle images and 1200 MESSIDOR images are used for testing in the first stage of the proposed system. In the second stage, the test Kaggle images from the first stage are used to train and test the second CNN where these images are divided randomly again into 80% for training and validation and 20% testing before feeding them into second stage. Table 3 shows the performance of the MESSIDOR dataset in terms of the 1R, 0.5R and 0.25R criteria for the two stages of the proposed system where on row one, we present the results of CCN1 for the test set of 1200 images (TS1M). Row two shows the results of CCN1 restricted to the images that are correctly detected within the 1R criterion (TS2M). Row three shows the TS2M set which is tested with CNN2 for comparison with row two. We can see that the results for these images are considerably improved by CNN2. Finally, on row four, we expand this test set to include incorrectly detected images (TS3M) from CCN1 demonstrating that, including these, the results remain strong and improved over the original idea of using CNN1 alone. Moreover, Table 4 presents the accuracy of the Kaggle dataset using the same criteria where on row one, we present the results of CCN1 for the test set of 3000 images (TS1M). Row two shows the results of CCN1 restricted to the images that are correctly detected within the 1R

criterion (TS2M). Row three shows the TS2M set tested on CNN2 for comparison with row two. We can see that the results for these images are considerably improved by CNN2. Finally, on row four, we expand this test set to include incorrectly detected images (TS3M) from CCN1.

**Table. 3** Performance (in terms of accuracy) of the network trained on Kaggle and tested on Messidor.

Model Name	Optic Disc			Fovea		
	1R	0.5R	0.25R	1R	0.5R	0.25R
CNN1+TS1M	97	86.3	47.5	96.6	76	35.3
CNN1+TS2M	100	88.9	49	100	78.8	36.5
CNN2+TS2M	100	97.9	86.2	100	94.6	69.2
CNN2+TS3M	<b>97</b>	<b>95</b>	<b>83.6</b>	<b>96.6</b>	<b>91.4</b>	<b>66.8</b>

**Table. 4** Performance (in terms of accuracy) of the network trained and tested on kaggle.

Model Name	Optic Disc			Fovea		
	1R	0.5R	0.25R	1R	0.5R	0.25R
CNN1+TS1K	96.7	87.4	51.9	95.6	83.4	51
CNN1+TS2K	100	90.1	55.6	100	87.9	54.3
CNN2+TS2K	100	99.1	93.4	100	94.9	73.3
CNN2+TS3K	<b>96.7</b>	<b>95.8</b>	<b>90.3</b>	<b>95.6</b>	<b>90.7</b>	<b>70.1</b>

The experimental results in **Table 3** and **Table 4** demonstrate that the proposed method can achieve accuracies in terms of the 1R criterion of 97% and 96.6% for detection of the OD and foveal centers respectively in MESSIDOR and 96.7% and 95.6% for the detection of the OD and foveal centers respectively in the Kaggle test set. On average, it only takes approximately 0.007 seconds to process a test image in both stages which is the fastest among all of the methods. Furthermore, the results show good performance when considering the 0.5R and 0.25R criteria. On the Kaggle test set, the obtained accuracies were 95.8% and 90.3% for OD detection for 0.5R and 0.25R respectively, while 90.7% and 70.1% were achieved for fovea detection in terms of these two criteria. On MESSIDOR, the detection accuracies were 95% and 83.6% for 0.5R and 0.25R for localizing the OD while the obtained accuracy results for the foveal center detection were 91.4% and 66.8% for the 0.5R and 0.25R criteria.

**Table 5** presents the results of our method and other methods reported in the literature. Detection accuracy, computational time, evaluation criterion and the dataset used are presented in this table for previous work where they are available in the original paper. **Fig. 8** and **Fig. 9** show some example detection results on the testing dataset. In **Fig. 8**, examples with accurate detections of the OD and fovea centers are presented while **Fig. 9** shows images with incorrect detections. **Fig.10** shows how the second stage CNN improves the detection performance over the first stage CNN.

## 5. DISCUSSION

A novel approach based on a multiscale sequential deep learning technique has been proposed for the simultaneous detection of the centers of the OD and fovea in color fundus

images. The designed CNNs achieve the detection by extracting complex data representations from retinal images without the need of human supervision. It has been demonstrated that the performance of our proposed system can outperform competing approaches.

It is worth mentioning that many different criteria were used by others in the literature to evaluate performance in detecting the OD and foveal centers when compared with the ground truth.

The Euclidean distance between the obtained OD and fovea center locations and their actual locations were often used as the evaluation measure. For example, many studies [12], [20]- [22] have established that the obtained detection of the OD (or foveal) center is correct if their Euclidean distances to the actual centers is within half the OD diameter (or one OD radius). This is the widely accepted 1R rule.

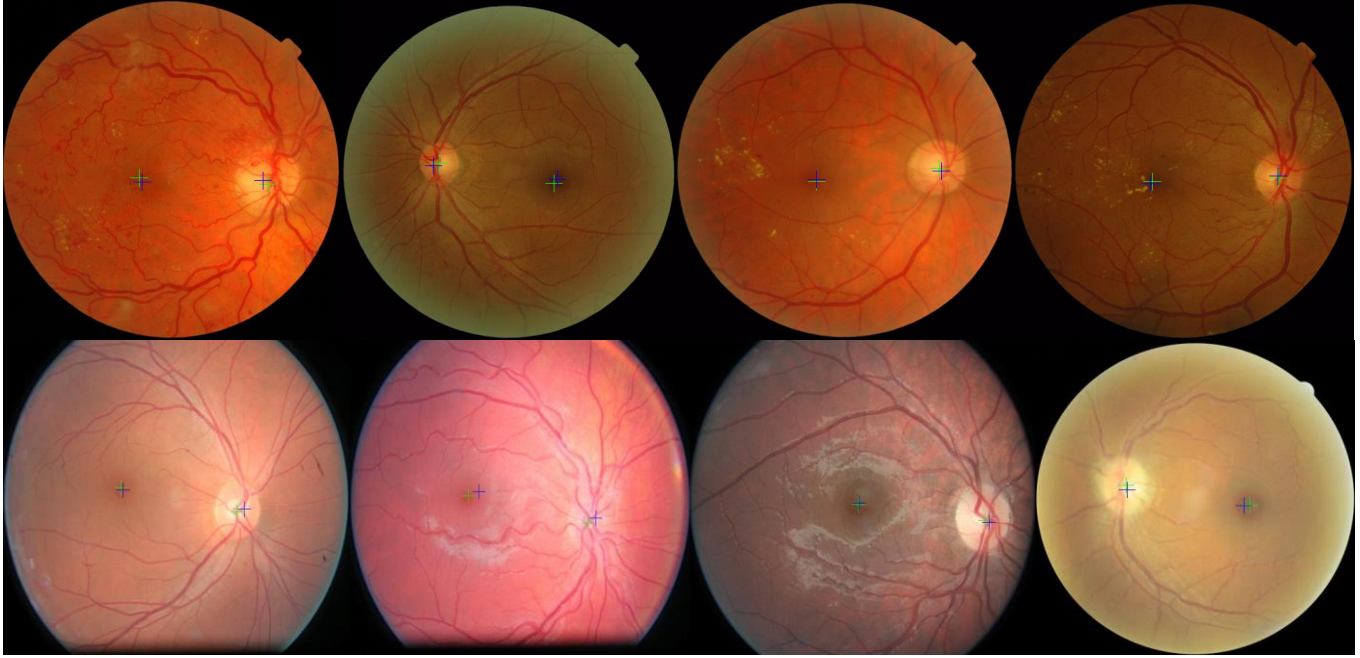
There is a problem in using the 1R rule for evaluation when the OD radius is not available. In order to alleviate this problem, Yu *et al.* [12] estimated the OD radius based on the field of view (FOV) of the retina and the image size. Three radii of 70, 100 and 110 pixels were used in correspondence to the three different sizes of the MESSIDOR images. Using this criterion, the authors detected the location of the OD correctly in 1189 out of the 1200 images in the MESSIDOR dataset. Following Yu's approach to estimate the OD radius, Giachetti *et al.* [20] reported an accuracy of 99.66% for OD detection and 99.1% for fovea detection and used the fast radial symmetry transform to achieve that. However, for the same MESSIDOR dataset Gegundez-Arias *et al.* [21] and Aquino *et al.* [22] used different OD radii in their study where the OD radii were fixed to 68, 103 and 109 pixels. Aquino *et al.* [22] reported an accuracy of 98.24% for the detection of the fovea. For this study, the 1R rule has been followed but the OD radius was defined by annotation results from experienced graders. As such, our rule should be more accurate. This has highlighted the issue that it is difficult to accurately compare detection performance between different methods as the criterion may be different.

The other issue for comparing results from different studies is that the number of images used were different. Even when studies used the same dataset, the way in which they used the dataset was not entirely clear. For instance, although Yu *et al.* [12] reported results on 1200 MESSIDOR images, they may have used the whole dataset in tailoring their detection method. This implies they have used the data to train their method and tested on the same dataset, which means their method may have overfit the data. Our study has split the Kaggle dataset into training and testing portions. Testing images of Kaggle have not been used until the network was trained using the separate training set. This suggests that our method should have better generalization ability. Furthermore, we use a completely unseen test set (MESSIDOR) to prove this generalization ability.

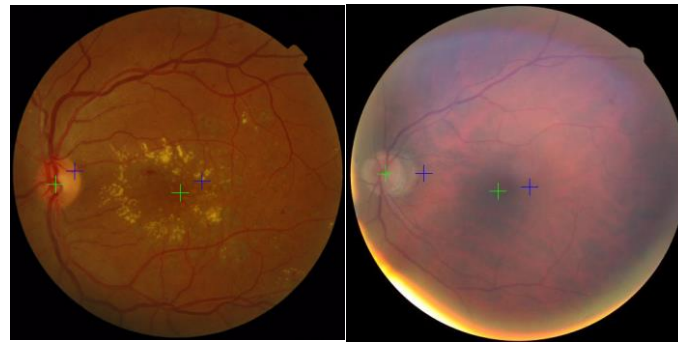
**Table. 5** Results of the proposed methodology for optic disc (OD) and fovea (F) detection compared with the existing methods in the literature

Authors	Approach	Detected landmarks	IR criterion (pixels)	Success rate	Running time	Dataset (Name, size, [images size])
Hoover [4]	Relationship between OD and blood vessels, fuzzy convergence algorithm	OD	60	Acc.: 89%	4 min.	(STARE [40], 81, [605×700])
Foracchia [5]	Geometrical model, blood vessels direction	OD	60	Acc.: 97.53%	2 min.	(STARE, 81, [605×700])
Li [15]	Parabola fitting	F	NA	Sensitivity: 100%	NA	(Local, 35, [512×512])
Fleming [6]	Visual characteristics of blood vessels, fovea, and OD	OD, F	119	Acc.: 98.4% , 96.5%	2 min.	(Local, 1056, [2160×1440])
Tobin [7]	Characteristics of blood vessels in relation to OD position	OD, F	65	Acc. : 90.4%, 92.5%	NA	(Local, 345, [1024×1152])
Niemeijer [17]	Cost function and a point distribution model	OD,F	50 50	Acc. : 98.4%, 94.4%, Acc.: 94% , 92%	10 min.	(Local, 500), (Local, 100), [768×576 - 2048×1536]
Aliaa Youssif [8]	2D Gaussian matched filter	OD	60 NA	Acc.: 98.77% Acc.: 100%	3.5 min.	(STARE, 81, [605×700]), (DRIVE [41], 40, [ 565 × 584])
Niemeijer [9]	k-NN regressor	OD, F	50 50	Acc.: 99.4%, 96.8% Acc.: 93%, 89%	7.6 sec.	(Local, 500), (Local, 100), [768×576 - 2048×1536]
Zhu [10]	Sobel operator, Hough transform	OD	40	Acc.: 90%	NA	(DRIVE, 40, [565×584])
Lu [11]	Circular transformation	OD	60 60 NA	Acc.: 99.75% Acc.: 97.5% Acc.: 98.77%	5 sec.	(STARE, 81, [605×700]), (ARIA [42], 120, [576×768]) (MESSIDOR, 1200, [1440×960, 2240×1488, 2304×1536])
Welfer [18]	Selection of ROI and morphology	F	34 34	Acc.: 100%, Acc.: 92.13%	NA	(DRIVE, 40, [565×584]), (DIARETDB1[43], 89, [640×480])
Yu [12]	Template matching technique	OD	70, 100 ,110	Acc.: 99%	4.7 sec.	(MESSIDOR, 1200, [1440×960, 2240×1488, 2304×1536])
Qureshi [19]	Combining the prediction of multiple algorithms	OD, F	NA	Acc.: 97.64% 96.79% Acc.: 97.79%,98.74 Acc.: 100%, 91.73%	NA	(DIARETDB0 [44], 130, [1500 ×1152]), (DIARETDB1, 89, [1500 ×1152]), (DRIVE, 40, [565 ×584])
Dehghani [13]	Template implemented from three histograms	OD	NA	Acc.: 100%, Acc.: 91.36% Acc.: 98.9%	27.6 sec.	(DRIVE, 40, [565×584]), (STARE, 81, [605×700]), (Local, 273, [720 ×576])
Giachetti [20]	Fast radial symmetry transform	OD, F	70, 100 ,110	Acc.: 99.66%, 99.1%	5 sec.	(MESSIDOR, 1200, [1440×960, 2240×1488, 2304×1536])
Gegundez-Arias [21]	Priori known anatomical features and thresholding	F	68, 103,109	Acc.: 96.92%	0.94 sec.	(MESSIDOR, 1200, [1440×960, 2240×1488, 2304×1536])
Aquino [22]	Visual and anatomical macula and OD feature-based method	F	68, 103,109 82	Acc.: 98.24% Acc.: 94.38%	10.88 sec.	(MESSIDOR, 1136, [1440×960, 2240×1488, 2304×1536]), (DIARETDB1, 89, [1500 ×1152]),
Harangi [14]	Ensemble-based framework (combining probability models)	OD	NA	Precision: 98.46% Precision: 98.88% Precision: 100% Precision: 98.33%	0.25 sec.	(DIARETDB0, 130, 1500×1152), (DIARETDB1, 130, 1500×1152), (DRIVE, 40, [565×584]), (MESSIDOR, 1200, [1440×960, 2240×1488, 2304×1536])
Proposed method	Deep neural network	OD, F	Variable	Acc.:97%, 96.6% Acc.:96.7%, 95.6%	0.007 sec.	MESSIDOR (1200) Kaggle





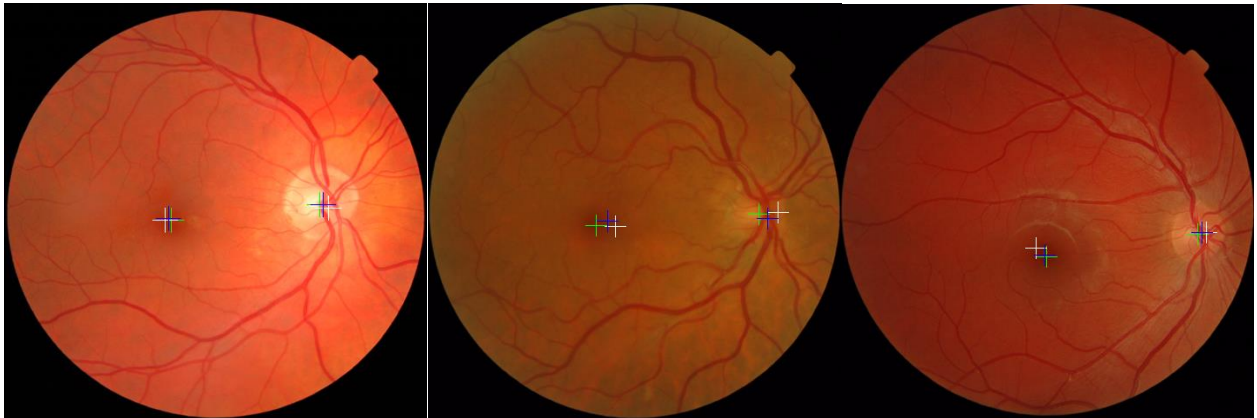
**Fig. 8** Examples of correct joint OD-Fovea detection results. First row examples from MESSIDOR and second row from Kaggle. The green plus signs refer to the locations annotated by ophthalmologists while the blue ones indicate the results of our proposed method.



(1)

(2)

**Fig. 9** Examples of incorrect OD and fovea detection results. (1) Incorrect detection from MESSIDOR; (2) Incorrect detection from Kaggle.



**Fig. 10** Examples of fundus images show the original centers (green plus), centers obtained from CNN1 (white plus), and centers from CNN2 (blue plus). It is clear that CNN2 improves the location accuracy.

Moreover, unlike most of the previous methods in the literature where only the 1R value has been reported, we report the accuracy based on the 0.5R and 0.25R criteria in addition to the 1R criterion. From the 0.5R and 0.25R reported accuracies, we notice that the performance has significantly improved by exploiting and analyzing the ROI for both OD and F in the second stage of the proposed system.

Although our network has provided competitive results, the network architecture may not be the optimal one as training the CNNs involve many hyperparameter settings such as regularization strength, the initial learning rate, and schedule of learning rate decay. Performing hyperparameter searches is considered a tricky and critical task [45]. Also, the number of convolution and pooling layers and the number and size of filters in each layer in CNNs are usually chosen empirically. As a result, the optimal network architecture and proper settings of these hyperparameters in the training stage are decided from experience and they are hard to find by non-expert humans [46]. In spite of these hyperparameter setting challenges in the training stage, once the network is trained, no expert is required to detect the landmarks in the test stage. Although data augmentation is considered to be useful in improving the performance of the CNNs, it is not clear what the best strategy is to achieve the best results. From our work, it is noted flipping horizontally is beneficial. However, rotation of images did not seem to improve the performance (results not shown).

## 6. CONCLUSION

We have demonstrated that our proposed method is capable of achieving excellent results in the detection of the optic disc and fovea in fundus images. One of the most important advantages of the proposed method is that it is less sensitive to preprocessing. It can be noticed that applying contrast enhancement as a preprocessing step improves the performance of the network, but not by very much. The current results were achieved without optimizing the parameters of the contrast enhancement method. Another advantage of our approach is that it does not necessitate the need of vessel segmentation or border localization in order to detect the OD and foveal centers. This will be useful when processing images of poor quality.

It has been proved that the ability to learn hierarchies of concepts, implementing multiple layers of abstraction in deep learning can be used for the detection landmarks in challenging medical applications. Likewise, the results of the proposed method suggest that deep learning can be used to address similar problems in other clinical applications such as screening and the diagnosis of diabetic retinopathy, age related macular degeneration and glaucoma. Moreover, as a result of the effectiveness of the deep neural network performance, this strategy will be investigated in our future work to grade the severity of retinal diseases such as diabetic retinopathy.

In conclusion, a new deep neural network approach has been proposed for the detection of the OD and foveal centers in color fundus images. Our proposed approach has produced promising results. This approach could be further developed and used as a crucial part of future automated diagnosis and grading software for the better management of eye disease.

## ACKNOWLEDGEMENT

We thank Mr D. G. Parry and Mrs S. Leach from the Liverpool Reading Center at St Paul's Eye Unit, Royal Liverpool University Hospital, for providing us with the ground truth annotations. B. Al-Bander was financially supported by the Higher Committee for Education Development in Iraq (Grant No. 182).

## REFERENCES

- Gagnon, L., Lalonde, M., Beaulieu, M., & Boucher, M. C. (2001, July). Procedure to detect anatomical structures in optical fundus images. In *Medical Imaging 2001* (pp. 1218-1225). International Society for Optics and Photonics.
- Larsen, H. W. (1976). *The ocular fundus: a color atlas*. WB Saunders Company.
- Schwiegerling, J. (2004, November). *Field guide to visual and ophthalmic optics*. Bellingham, WA, USA: Spie.
- Hoover, A., & Goldbaum, M. (2003). Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *Medical Imaging, IEEE Transactions on*, 22(8), 951-958.
- Foracchia, M., Grisan, E., & Ruggeri, A. (2004). Detection of optic disc in retinal images by means of a geometrical model of vessel structure. *Medical Imaging, IEEE Transactions on*, 23(10), 1189-1195.
- Fleming, A. D., Goatman, K. A., Philip, S., Olson, J. A., & Sharp, P. F. (2006). Automatic detection of retinal anatomy to assist diabetic retinopathy screening. *Physics in Medicine and Biology*, 52(2), 331.
- Tobin, K. W., Chaum, E., Govindasamy, V. P., & Karnowski, T. P. (2007). Detection of anatomic structures in human retinal imagery. *Medical Imaging, IEEE Transactions on*, 26(12), 1729-1739.
- Youssif, A. A. H. A. R., Ghalwash, A. Z., & Ghoneim, A. A. S. A. R. (2008). Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter. *Medical Imaging, IEEE Transactions on*, 27(1), 11-18.
- Niemeijer, M., Abramoff, M. D., & Van Ginneken, B. (2009). Fast detection of the optic disc and fovea in color fundus photographs. *Medical Image Analysis*, 13(6), 859-870.
- Zhu, X., Rangayyan, R. M., & Ells, A. L. (2010). Detection of the optic nerve head in fundus images of the retina using the Hough transform for circles. *Journal of Digital Imaging*, 23(3), 332-341.
- Lu, S. (2011). Accurate and efficient optic disc detection and segmentation by a circular transformation. *Medical Imaging, IEEE Transactions on*, 30(12), 2126-2133.
- Yu, H., Barriga, E. S., Agurto, C., Echegaray, S., Pattichis, M. S., Bauman, W., & Soliz, P. (2012). Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets. *Information Technology in Biomedicine, IEEE Transactions on*, 16(4), 644-657.
- Dehghani, A., Moghaddam, H. A., & Moin, M. S. (2012). Optic disc localization in retinal images using histogram matching. *EURASIP Journal on Image and Video Processing*, 2012(1), 1-11.
- Harangi, B., & Hajdu, A. (2015). Detection of the optic disc in fundus images by combining probability models. *Computers in Biology and Medicine*, 65, 10-24.
- Calimeri, F., Marzullo, A., Stamile, C., & Terracina, G. (2016, November). Optic Disc Detection Using Fine Tuned Convolutional Neural Networks. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on* (pp. 69-75). IEEE.
- Li, H., & Chutatape, O. (2004). Automated feature extraction in color retinal images by a model based approach. *Biomedical Engineering, IEEE Transactions on*, 51(2), 246-254.
- Niemeijer, M., Abramoff, M. D., & Van Ginneken, B. (2007). Segmentation of the optic disc, macula and vascular arch in fundus photographs. *Medical Imaging, IEEE Transactions on*, 26(1), 116-127.
- Welfer, D., Scharcanski, J., & Marinho, D. R. (2011). Fovea center detection based on the retina anatomy and mathematical morphology. *Computer Methods and Programs in Biomedicine*, 104(3), 397-409.
- Qureshi, R. J., Kovacs, L., Harangi, B., Nagy, B., Peto, T., & Hajdu, A. (2012). Combining algorithms for automatic detection of optic disc and

- macula in fundus images. *Computer Vision and Image Understanding*, 116(1), 138-145.
20. Giachetti, A., Ballerini, L., Trucco, E., & Wilson, P. J. (2013). The use of radial symmetry to localize retinal landmarks. *Computerized Medical Imaging and Graphics*, 37(5), 369-376.
  21. Gegundez-Arias, M. E., Marin, D., Bravo, J. M., & Suero, A. (2013). Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques. *Computerized Medical Imaging and Graphics*, 37(5), 386-393.
  22. Aquino, A. (2014). Establishing the macular grading grid by means of fovea center detection using anatomical-based and visual-based features. *Computers in Biology and Medicine*, 55, 61-73.
  23. Decenciere, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., & Charton, B. (2014). Feedback on a publicly distributed image database: the Messidor database. *Image Analysis and Stereology*, 231-234.
  24. Diabetic Retinopathy detection (2016, October 18). Retrieved from <https://www.kaggle.com/c/diabetic-retinopathy-detection>
  25. Zuiderveld, K. (1994, August). Contrast limited adaptive histogram equalization. In *Graphics Gems IV* (pp. 474-485). Academic Press Professional, Inc.
  26. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105).
  27. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580-587).
  28. Long, J. L., Zhang, N., & Darrell, T. (2014). Do convnets learn correspondence? In *Advances in Neural Information Processing Systems* (pp. 1601-1609).
  29. Ciresan, D., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in Neural Information Processing Systems* (pp. 2843-2851).
  30. Cernazanu-Glavan, C., & Holban, S. (2013). Segmentation of bone structure in X-ray images using convolutional neural network. *Adv. Electr. Comput. Eng.*, 13(1), 87-94.
  31. Li, S., & Chan, A. B. (2014). 3d human pose estimation from monocular images with deep convolutional neural network. In *Computer Vision--ACCV 2014* (pp. 332-347). Springer International Publishing.
  32. Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 34-42).
  33. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
  34. Ranzato, M. A., Huang, F. J., Boureau, Y. L., & LeCun, Y. (2007). Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on* (pp. 1-8). IEEE.
  35. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929-1958.
  36. Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In *International Conference on Artificial Intelligence and Statistics* (pp. 315-323).
  37. Lasagne/Lasagne. (2016, October 18). Retrieved from <https://github.com/Lasagne/Lasagne>
  38. Theano/Theano. (2016, October 18). Retrieved from <https://github.com/Theano/Theano>
  39. Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *International Conference on Artificial Intelligence and Statistics* (pp. 249-256).
  40. Hoover, A., Kouznetsova, V., & Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *Medical Imaging, IEEE Transactions on*, 19(3), 203-210.
  41. Staal, J., Abramoff, M. D., Niemeijer, M., Viergever, M. A., & Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *Medical Imaging, IEEE Transactions on*, 23(4), 501-509.
  42. Zheng, Y., Hijazi, M. H. A., & Coenen, F. (2012). Automated "disease/no disease" grading of age-related macular degeneration by an image mining approach. *Investigative Ophthalmology & Visual Science*, 53(13), 8310-8318.
  43. Kälviäinen, R. V. J. P. H., & Uusitalo, H. (2007). DIARETDB1 diabetic retinopathy database and evaluation protocol. *Medical Image Understanding and Analysis* 2007, 61.
  44. Kauppi, T., Kalesnykiene, V., Kamarainen, J. K., Lensu, L., Sorri, I., Uusitalo, H., & Pietilä, J. (2006). DIARETDB0: Evaluation database and methodology for diabetic retinopathy algorithms. *Machine Vision and Pattern Recognition Research Group, Lappeenranta University of Technology, Finland*.
  45. Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, & Adams, R. P. (2015). Scalable Bayesian optimization using deep neural networks. *arXiv preprint arXiv:1502.05700*.
  46. Hinton, G. E., Montavon, G., Orr, G., & Müller, K. R. (2012). Neural networks: tricks of the trade (pp. 599-619). *Lecture Notes in Computer Science*. Springer Berlin Heidelberg.